ERDC/EL TN-ECMI-04-1
November 2004

Classification of Land-Cover Types for the Fort Benning Ecoregion Using Enhanced Thematic Mapper Data: January 2003 Imagery

by Sam S. Jackson, Scott G. Bourne, Mark R. Graves

PURPOSE: Information regarding regional land cover is a fundamental requirement to support the long-term baseline ecosystem monitoring plan under the Strategic Environmental Research and Development Program (SERDP), Ecosystem Management Project (SEMP), Ecosystem Characterization and Monitoring Initiative (ECMI). The land cover characterization phase of this plan provides the foundation needed to derive vegetation density indices and land cover patterns. These characteristics are the primary visible expressions of the underlying ecosystem structure, function, and process at all spatial scales (Kress 2000). To meet the requirement for land cover information, Landsat 7 Enhanced Thematic Mapper (ETM) data were used to classify land cover types for the Fort Benning ecoregion. This technical note describes the procedures used to extract land cover information from the satellite imagery.

BACKGROUND: At a regional scale, land cover significantly affects biophysical factors such as surface albedo and sensible heat flux, and plays an important role in material cycling. Developing an accurate land cover classification is vital, since other landscape characteristics are directly linked to it.

Satellite imagery has been used since the 1970's as an accurate and cost-effective tool for deriving regional vegetation and land cover information. Digital processing techniques involving the statistical analysis of image data representing various portions of the electromagnetic spectrum allow definition of areas that reflect solar radiation in a like manner (the thermal band was not used in this analysis). These areas may then be related to land cover or vegetation types through the use of ground-truth data collected in the field.

Landsat 7 ETM data were selected for this study due to the spectral and spatial characteristics of the sensor, which have been documented as appropriate data for mapping broad vegetative types such as deciduous and evergreen forests (Schriever and Congalton 1993). In addition, Landsat 7 data are relatively inexpensive and the scenes required to cover the study area could be acquired quickly. The Landsat 7 satellite ETM sensor provides six spectral bands of imagery, each with a spatial resolution of 28.5 m. The ETM sensor also provides one panchromatic (black and white) band with 15-m spatial resolution and one thermal band with 60-m spatial resolution.

ETM data collected on January 24, 2003 (scene 7019037000302450-path 019/row 037 and scene 7019038000302450-path 019/row 038) were used in the study. Unfortunately, no acceptable leaf-on scenes were available during the necessary time frame to allow an optimum discrimination of vegetation types.

STUDY AREA: Fort Benning, GA, is located in west central Georgia, south of the city of Columbus, GA, and east of Phenix City, AL. The study area is comprised of approximately 181,395 ha, which includes the military base and a portion of the surrounding U.S. Geological Survey (USGS) Hydrologic

Unit Code (HUC) 03130003. The base covers 73,812 ha. Figure 1 illustrates the location of the military installation and the HUC. The land cover of the study area is dominated by a variety of evergreen and deciduous forest types distributed over moderate rolling topography.

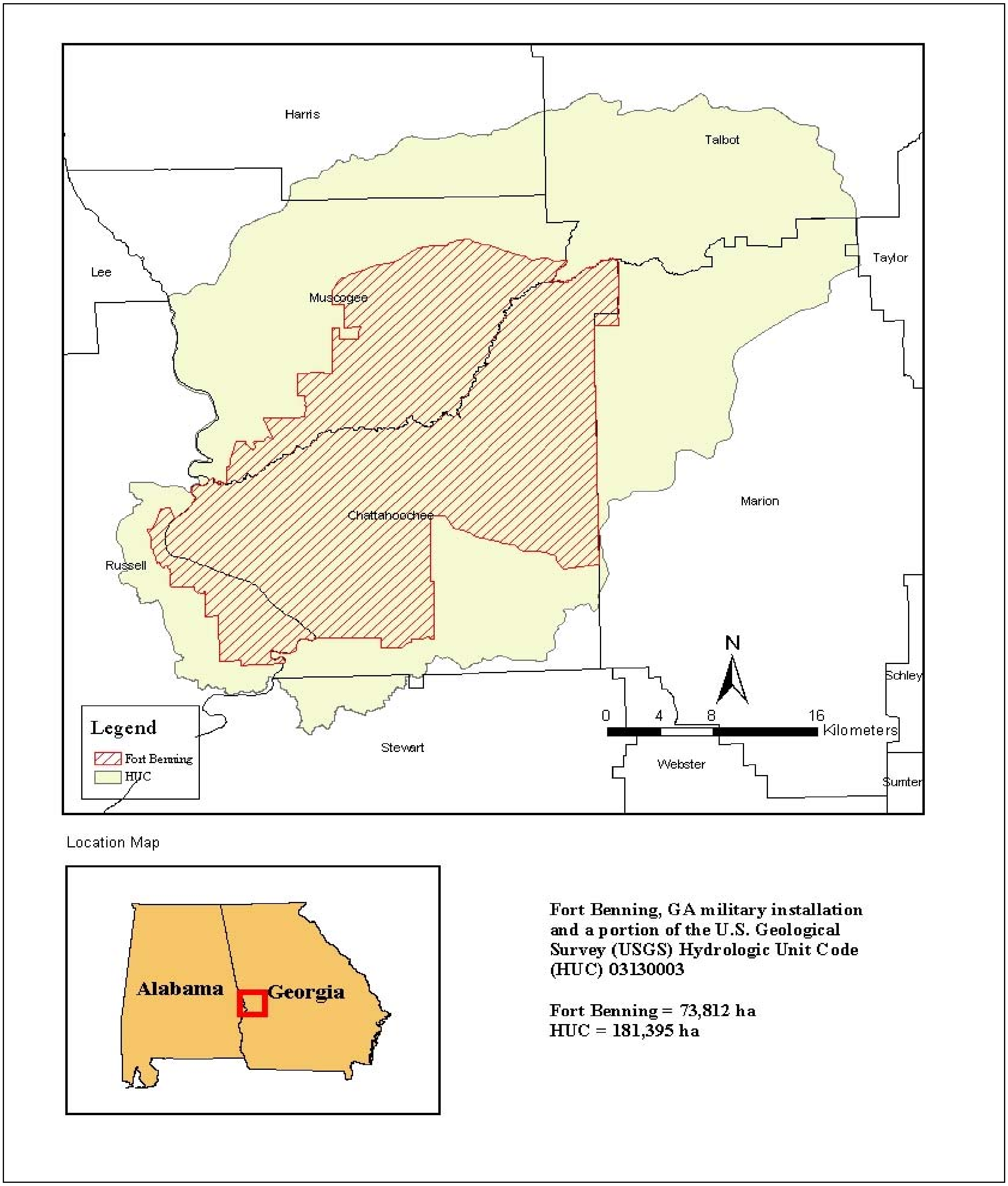

Figure 1. Location of study area

METHOD: ERDAS Imagine (Version 8.7) software was used to perform all image processing functions required to complete the land cover classification. A hybrid supervised/unsupervised classification methodology was used. In a supervised classification, the image analyst is responsible for defining "training areas," which represent all of the cover types he or she wishes to extract from an image. The training areas are used to extract the digital values from the imagery to produce "signatures" or statistical definitions of each cover class. However, it is often difficult to account for all the cover types in an image as well as variability within cover types. Unsupervised classification differs from supervised classification in that the computer (rather than the user) develops the signatures that will be used to classify the scene. The classification process results in a number of spectral classes, which the analyst must then assign (a posteriori) to information classes of interest. This requires knowledge of the terrain present in the scene as well as its spectral characteristics.

Fleming, Berkebile, and Hoffer (1975) outlined a hybrid approach for image classification that makes use of the benefits of both supervised and unsupervised approaches. This approach requires the following four steps:

1) Use clustering algorithms to determine the spectral classes into which the image can be divided.
2) Use ground truth data to assign information classes to the statistical clusters.
3) Use statistical distance measures to evaluate the initial clusters. Delete or merge clusters as necessary.
4) Using a maximum-likelihood algorithm, classify the entire image into the set of spectral classes.

To generate a better set of class statistics for the bare ground land cover type, cantonment areas and paved roads were removed from the Landsat digital image before the unsupervised clustering algorithm was used. These land cover types have been mapped and are available in the Fort Benning Geographical Information System (GIS) database. This would produce a better training class for bare ground by eliminating any chance of confusing bare ground with paved roads or features in the cantonment area that have similar reflectance properties.

Definition of Initial Clusters: In unsupervised classifications, statistical clustering algorithms are used to analyze the digital values in each band of imagery and to determine the number of statistically distinct features (clusters) in the image. In this study, an unsupervised iterative self-organizing data analysis (ISODATA) clustering algorithm was used. ISODATA is a widely used clustering algorithm that makes a large number of passes through an image using a minimum spectral distance formula to form clusters. It begins with either arbitrary cluster means or means of an existing signature set, and each time the clustering repeats, the means of these clusters are shifted. The new cluster means are used for the next iteration. This iterative process continues until statistically distinct features emerge.

To perform ISODATA clustering, it was necessary to specify the Landsat bands to be used for the classification. For this study the three visible bands (TM1, TM2, and TM3), one near infrared band (TM4), and two middle infrared bands (TM5 and TM7) were used. Next, the maximum numbers of clusters were determined. Fifty initial clusters were requested. Such a large number of clusters were requested to provide a wide variation of land cover types that could be easily identified. Requesting too few clusters could have caused the ISODATA clustering method to combine different land cover types. Since each cluster is the basis for a class, this number becomes the maximum number of classes to be formed.

The ISODATA process began by determining the 50 arbitrary cluster means that were requested. On the first iteration of the ISODATA algorithm, the means of the 50 clusters were arbitrarily determined. After each iteration a new mean for each cluster was calculated based on the actual spectral locations of the pixels in the cluster, instead of the initial arbitrary calculation. These new means were used for defining clusters in the next iteration. The process continued until there was little change between iterations. The convergence threshold was set to 95 percent, which is the maximum percentage of pixels whose class values are allowed to be unchanged between iterations. After each iteration the normalized percentage of pixels whose assignments are unchanged since the last iteration is displayed, and when the percentage of unchanged pixels reaches 95 percent, the classification is completed (Smith, Pyden, and Cole 1995).

Evaluation of Clusters: Statistical separability tools were used to determine if clusters from ISODATA were to be used as a class in the final classification, combined with another class generating a new land cover type, or discarded. Signature separability is a statistical measure of distance between two signatures. Separability can be calculated for any combination of bands that will be used in the classification generated by the ISODATA clustering algorithm. Bands 4 and 5 were used to calculate the separability of the class signatures.

Transformed divergence (TD), used in this study, is a measure of statistical separation between category response patterns and uses a covariance-weighted distance between category means to determine whether class signatures are separable. TD values have an upper (2000) and a lower bound (0). If the calculated divergence is equal to the upper bound, then the signatures can be said to be totally separable in the bands being analyzed. A calculated divergence of zero means that the signatures are inseparable. For this study, any class combination with a TD above 1500 was considered separable. Any class combination with a TD below 1500 was combined or one of the classes was disregarded. The next step was to select the class combination with the lowest TD. Each class was analyzed individually to identify the class type (i.e., hardwood forest, bare ground, etc.) through the use of high-resolution digital ortho photographs. After the class types were determined, the TD separability values were used to decide which class types should be deleted or combined. For example, signature 5 and signature 8 had a TD separability of 1036. Signature 8 had separability values lower than 1500 with two other class signatures, while signature 5 was inseparable from only one other class signature. Signature 5 was considered a better class and signature 8 was therefore deleted. An analysis of the number of pixels in each class and the distribution of pixels in the classified images are accomplished before generating the final classification.

After analyzing class combinations with a TD of less than 1500, the final number of separable classes was 32. Each of the final 32 classes was found to belong to one of the following 8 classes in Table 1 (there was more than one statistically separable type of hardwood, for example).

Table 1
Land Cover Types

Water	Evergreen/Hardwood
Evergreen Planted	Scrub/Shrub
Evergreen	Herbaceous
Hardwood	Bare Ground

Maximum-Likelihood Classification: The next stage of the classification process involved using a maximum-likelihood algorithm. The 32 final class signatures developed using ISODATA were input to the maximum-likelihood algorithm. The maximum-likelihood classification makes use of the statistical parameters developed through the ISODATA process and, in addition, uses estimates of probability distributions to determine the relative likelihood that a pixel belongs to a certain class.

Even though these classes had been identified once using ISODATA, each class was checked with Digital Orthophoto Quads (DOQ's) and Forest Stand coverage to make sure that classes were mapped correctly.

Refinement of Final Product: Once the maximum-likelihood classification was complete, many isolated, very small classes were present. A filtering approach was used to eliminate the small areas that occurred through the classified image. These small areas are 1 to 9 pixels in size. A 3×3 low-pass filter was used to eliminate these areas. The first application of the filter removes any small areas in the classified image. The second application eliminates any small areas generated by using the filter the first time.

Using the capabilities of the Imagine software, layers in the Fort Benning GIS database representing cantonment areas, paved roads, and tank trails were merged onto the final classification. In addition, the urban delineation was performed through the use of automated feature extraction software, specifically Visual Learning System's Feature Analyst (Version 3.4), and the results were clipped to the HUC boundary and merged onto the final classification (Visual Learning Systems 2002).

RESULTS: During the classification process, 50 training classes were initially generated from the Landsat image. Classes with spectral similarities were then aggregated to derive a land cover type map with 32 classes. The 32 classes were then further combined to form a final land cover map with 8 unique classes. The land cover classification is displayed in Figure 2. Final classification results for the entire HUC and the military installation are summarized in Tables 2 and 3, respectively.

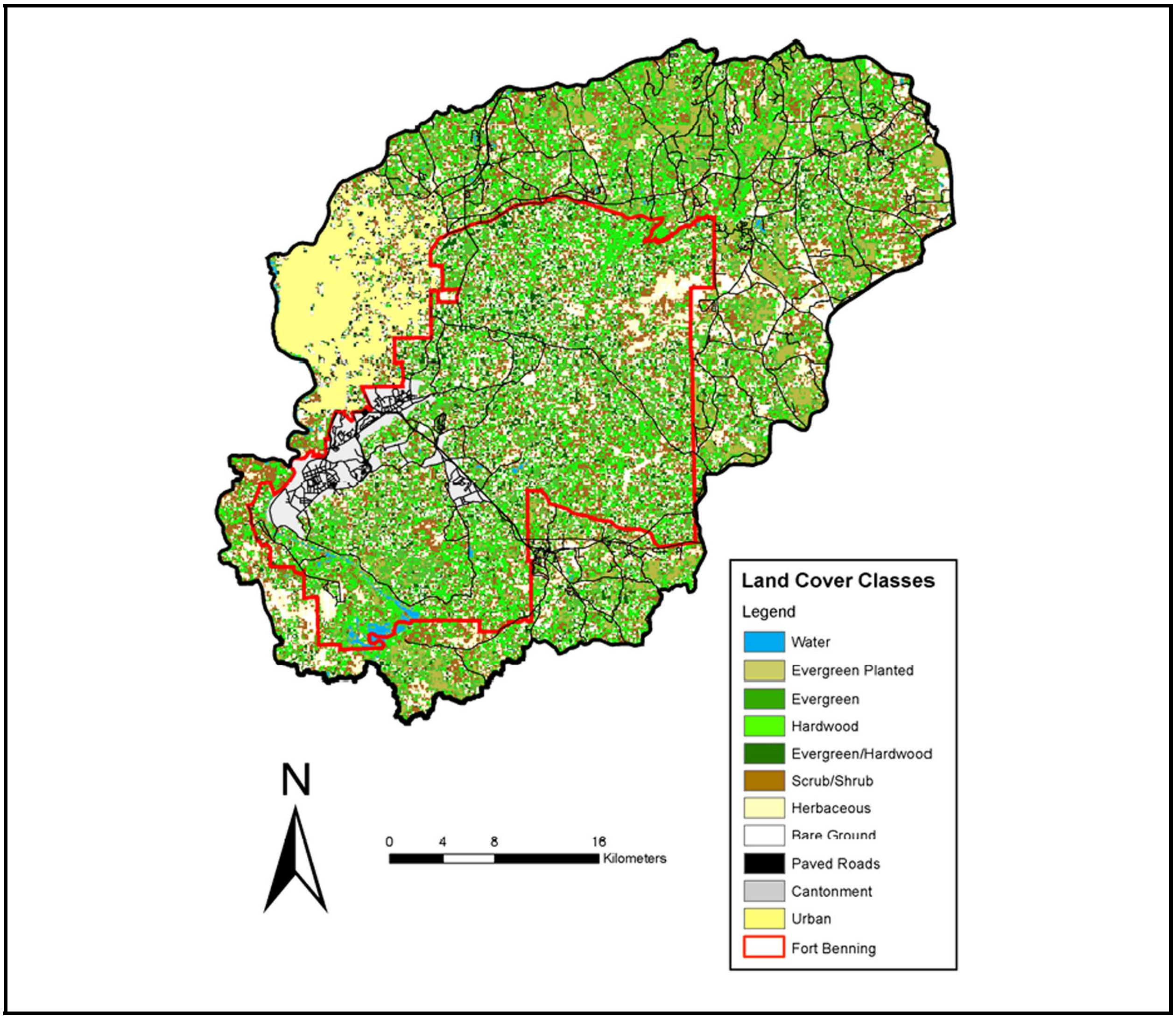

Figure 2. Land cover classification for Fort Benning Ecoregion

Table 2
Unsupervised Classification Results of the HUC

Name	Area in Hectares	Area in Acres	% of Total
Water	2,587	6,393	1.4%
Evergreen Planted	14,091	34,819	7.8%
Evergreen	30,688	75,831	16.9%
Hardwood	48,105	118,869	26.5%
Evergreen/Hardwood	29,598	73,138	16.3%
Scrub/Shrub	18,705	46,222	10.3%
Herbaceous	12,840	31,728	7.1%
Bare Ground	4,926	12,174	2.7%
Paved Roads	2,883	7,124	1.6%
Cantonment	5,426	13,408	3.0%
Urban	11,753	29,041	6.5%
Total	**181,602**	**448,748**	**100%**

Table 3
Unsupervised Classification Results of the Fort Benning Military Installation

Name	Area in Hectares	Area in Acres	% of Total
Water	1,031	2,547	1.4%
Evergreen Planted	1,988	4,913	2.7%
Evergreen	14,893	36,800	20.1%
Hardwood	22,023	54,420	29.8%
Evergreen/Hardwood	17,343	42,857	23.5%
Scrub/Shrub	5,759	14,231	7.8%
Herbaceous	3,302	8,158	4.5%
Bare Ground	1,392	3,439	1.9%
Paved Roads	766	1,893	1.0%
Cantonment	5,426	13,408	7.3%
Total	**73,923**	**182,667**	**100%**

Classification Accuracy Assessment: As with all land cover maps produced using this process, the result of the classification represents only a generalization of the real landscape types. Therefore, it is necessary to check the accuracy of the land cover classification with ground truth data, if available (Ahmad, Jupp, and Nunez 1992). Ground truth data and the classification results must be compared and statistically analyzed. Contingency tables or error matrices are common statistical methods used to compare the results obtained from the classification. An error matrix analysis provides a natural framework for the convenient display of results that can also be used for analysis. This is an effective tool that presents the overall accuracy of the classification as well as the producer and user accuracy of each category (Congalton, Oderwald, and Mead 1983).

The classification accuracy assessment was based on visual interpretation using very high quality 0.5-m resolution airborne digital multispectral imagery as reference data. The imagery was acquired in November 2003. A stratified random sample design was used to randomly allocate reference points throughout each land cover class. Point coordinates generated from the final classified Landsat image were cross-referenced with the same point in the high-resolution airborne MS image and their accuracy determined. Since the result of each class comparison has only one of two outcomes, correct or incorrect, a binomial probability distribution (Equation 1) was used to calculate the random sample size (Environmental Systems Research Institute, Inc., National Center for Geographic Information and Analysis, and the Nature Conservancy 1994):

$$N = \frac{Z^2 p q}{E^2} \tag{1}$$

where:

N = number of samples

Z = generalized from the standard normal deviate, 1.96 for 95% two-sided confidence level, 2

p = required percent accuracy, 90%

q = 100 - p

E = allowable error (standard deviation from the mean), 10%

Van Genderen and Lock (1977) demonstrated that a minimum sample size of 20 per class is required for a classification accuracy of 85 percent, while 30 observations per class are required for 90-percent accuracy (at the 0.05 significance level). Using the stated values above as input into the binomial distribution equation, the sample size required for this study was found to be 36 per class. The level of agreement or disagreement between the Landsat classification and the reference data is shown in Table 4. The overall accuracy of the classification is 86.1 percent. The overall accuracy was calculated by summing the main diagonal elements of the error matrix and dividing by the total number of samples (N = 288). The four major forest stand classes of interest including evergreen/planted, evergreen, hardwood, and evergreen/hardwood mix revealed user accuracies of 94, 81, 77, and 78 percent, respectively. Data were not available to check the accuracy of the paved roads, cantonment, or urban land cover types.

CONCLUSION: The land cover classification technique discussed in this technical note remains a valid method of discriminating broad land cover types from Landsat 7 ETM Imagery for the Fort Benning Ecoregion. However, it should be noted that extracting detailed urban features through the use of Feature Analyst, as performed in this analysis, might be more applicable with higher resolution imagery than coarse-resolution Landsat 7 ETM. In addition, further considerations should be directed towards evaluating and improving accuracy assessment procedures with the inclusion of possible ground-truth data to supplement the approach used in this analysis.

ACKNOWLEDGMENTS: This report was prepared for the Ecosystem Characterization and Monitoring Initiative (ECMI) sponsored by the Strategic Environmental Research and Development Program (SERDP) Ecosystem Management Project (SEMP). The technical monitor was Dr. Robert Holst, SERDP Program Manager.

Table 4
Agreement Between Landsat Classification and Reference Data

Reference Data: Classification	Water	Evergreen Planted	Evergreen	Hardwood	Evergreen/Hardwood	Scrub/Shrub	Herbaceous	Bare Ground	Row Total	Producer's Accuracy %
Water	***36***	0	0	0	0	0	0	0	36	100
Evergreen Planted	0	***30***	6	0	0	0	0	0	36	83
Evergreen	0	0	***35***	0	1	0	0	0	36	97
Hardwood	0	0	0	***30***	5	1	0	0	36	83
Evergreen/Hardwood	0	1	0	7	***28***	0	0	0	36	78
Scrub/Shrub	0	0	0	2	0	***33***	1	0	36	92
Herbaceous	0	0	2	0	0	7	***27***	0	36	75
Bare Ground	0	1	0	0	2	1	3	***29***	36	80
Column Total	36	32	43	39	36	42	31	29	***288***	
User's Accuracy %	100	94	81	77	78	78	87	100		

Overall Accuracy = 86.1%

The work was performed under the direction of the Ecosystem Evaluation and Engineering Division (EE) of the Environmental Laboratory (EL), U.S. Army Engineer Research and Development Center (ERDC). The EL Principal Investigator was Mr. Sam S. Jackson and co-investigators were Mr. Scott G. Bourne and Mr. Mark R. Graves, EL. Project Manager for the ECMI was Dr. Rose Kress, EL, and Program Manager for the SEMP was Mr. William Goran of the Construction Engineering Research Laboratory (CERL), ERDC, Champaign, IL.

Many individuals contributed to the support of this project, including the following: Mr. John Brent, Mr. Pete Swiderek, and Mr. Rusty Bufford of Fort Benning, GA, the host site for the SEMP; Drs. Rose Kress and David Price of EL, and Ms. Elizabeth Lord, Dyntel Corporation.

At the time of publication, Director of EL was Dr. Ed Theriot and Chief of EE was Dr. David J. Tazik. Dr. James R. Houston was Director of ERDC, and COL James R. Rowan, EN, was Commander.

POINTS OF CONTACT: For additional information contact Mr. Sam S. Jackson, U. S. Army Engineer Research and Development Center (ERDC), Waterways Experiment Station, Vicksburg, MS (601-634-

3317, *Sam.S.Jackson@erdc.usace.army.mil*). This technical note was written by Mr. Jackson, Mr. Scott G. Bourne, and Mr. Mark R. Graves, ERDC. This document should be cited as follows:

> Jackson, S. S., Bourne, S. G., and Graves, M. R. (2004). "Classification of land-cover types for the Fort Benning ecoregion using enhanced thematic mapper data: January 2003 imagery," *SERDP Technical Notes Collection,* ERDC/EL TN-ECMI-04-1, U.S. Army Engineer Research and Development Center, Vicksburg, MS.

REFERENCES

Ahmad, W., Jupp, L.B., and Nunez, M. (1992). "Land cover mapping in a rugged terrain area using Landsat MSS data," *Int. J. Remote Sensing* 13(4), 673-683.

Congalton, R., Oderwald, R. and Mead, R. (1983). "Assessing Landsat classification accuracy using discrete multivariate statistical techniques," *Photogrammetric Engineering and Remote Sensing* 49(12): 1671-1678.

Environmental Systems Research Institute, Inc., National Center for Geographic Information and Analysis, and The Nature Conservancy. (1994). "Accuracy assessment procedures," National Park Service/National Biological Survey (NBS/NPS) Vegetation Mapping Program, Washington, DC.

Fleming, M. D., Berkebile, J. S., and Hoffer, R. M. (1975). "Computer aided analysis of Landsat-1 MSS data: A comparison of three approaches including a modified clustering approach," Information Note 072475, Laboratory for Application of Remote Sensing, West Lafayette, IN.

Kress, M. R. (2000). "Executive Summary, Long-Term Monitoring Program, Fort Benning, GA, Version 2.1," Technical Report ERDC/EL TR-0l-15, U.S. Army Engineer Research and Development Center, Vicksburg, MS.

Schriever, J. R., and Congalton, R. G. (1993). "Mapping forest cover-types in New Hampshire using multi-temporal Landsat Thematic Mapper data," *ASPRS/ACSM Annual Convention & Exposition, 15-19 February (New Orleans, Louisiana),* 3:333-342.

Smith, C., Pyden, N., and Cole, P. (1995). *ERDAS Field Guide.* 3rd ed., ERDAS, Inc., Atlanta, GA.

Van Genderen, J. L., and Lock, B. F. (1977). "Testing Land-Use Map Accuracy," *Photogrammetric Engineering and Remote Sensing* 43(9): 1135-1137.

Visual Learning Systems. (2002). *User Manual, Feature Analyst Extension for ArcView 8.1*, Visual Learning Systems, Inc., Missoula, MT.